

AD\_\_\_\_\_

Award Number: W81XWH-11-1-0304

**TITLE:** Avoiding microRNA Function Through Alternative Polyadenylation in Prostate Cancer

**PRINCIPAL INVESTIGATOR:** Eric J. Wagner, Ph.D.

**CONTRACTING ORGANIZATION:** The University of Texas Health Science Center at Houston  
Houston, TX 77030

**REPORT DATE:** April 2012

**TYPE OF REPORT:** Final

**PREPARED FOR:** U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

**DISTRIBUTION STATEMENT:** Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE 01-04-2012	2. REPORT TYPE Final	3. DATES COVERED 1 apr 2011 - 31 mar 2012		
4. TITLE AND SUBTITLE  Avoiding microRNA Function Through Alternative Polyadenylation in Prostate Cancer		5a. CONTRACT NUMBER		
		5b. GRANT NUMBER W81XWH-11-1-0304		
		5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S)  Eric J Wagner  E-Mail: Eric.J.Wagner@uth.tmc.edu		5d. PROJECT NUMBER		
		5e. TASK NUMBER		
		5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  The University of Texas Health Science Center at Houston Houston, TX 77030-5400		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)  U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012		10. SPONSOR/MONITOR'S ACRONYM(S)		
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited				
13. SUPPLEMENTARY NOTES				
14. ABSTRACT  Alternative Cleavage and Polyadenylation (APA) is an exciting, yet poorly contributor to tumorigenesis. The current questions that require immediate question are what protein(s) regulate this process and what are the genes most critical that get regulated by this event. Here we tested the hypothesis that APA occurs during acquisition of androgen-independence and the H3K36 trimethylase, Setd2 may modulate this event. While we did not detect any effects to APA due to Setd2, we did determine that the RNA binding protein, Cflm25, is dramatically downregulated in androgen receptor negative prostate cancer cells and that this protein is involved in regulating APA. We used RNA-seq to identify Cflm25 targets and found a significant enrichment in regulated genes that are involved in androgen signaling. These data uncover a previously unknown form of regulation on androgen signaling and give rise to a novel research platform to base future, more substantial research.				
15. SUBJECT TERMS Alternative cleavage and polyadenylation, RNA-seq, androgen signaling				
16. SECURITY CLASSIFICATION OF:  a. REPORT U		17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 21	19a. NAME OF RESPONSIBLE PERSON USAMRMC
b. ABSTRACT U				19b. TELEPHONE NUMBER (include area code)
c. THIS PAGE U				

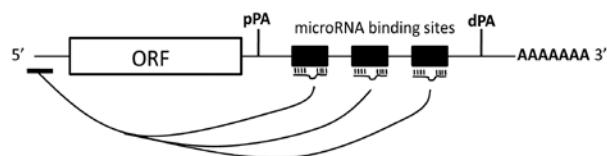
## Table of Contents

	<u>Page</u>
<b>Introduction.....</b>	<b>1</b>
<b>Body.....</b>	<b>2</b>
<b>Key Research Accomplishments.....</b>	<b>5</b>
<b>Reportable Outcomes.....</b>	<b>5</b>
<b>Conclusion.....</b>	<b>6</b>
<b>References.....</b>	<b>6</b>
<b>Appendices.....</b>	

## Introduction.

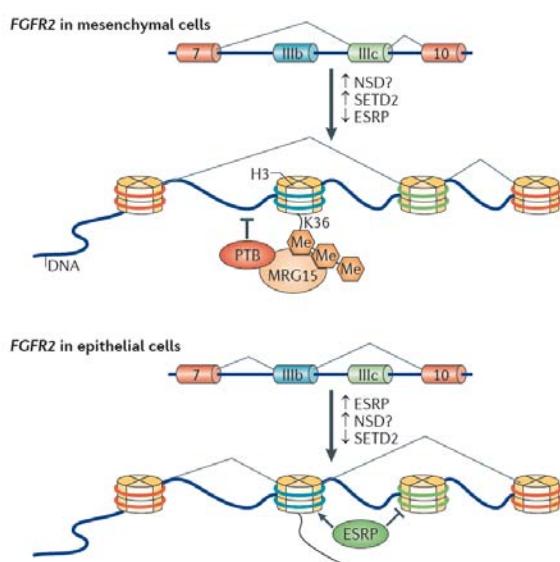
There were *two primary goals* of this one-year Hypothesis Development Award: Identify genes whose alternative cleavage and polyadenylation (APA) may be altered during the acquisition of androgen-independence and to determine if H3K36 methylation changes regulate this event. This document will describe the results of our efforts and what we conclude from these results. As shown below, we found no evidence for a role for H3K36 methylation changes in the regulation of APA but did discover a key factor that does regulate APA and its expression is reduced in androgen receptor negative cells. Moreover, we have conducted RNA-Seq to identify targets of this factor and found androgen signaling to be a major pathway affected upon depletion of this protein. In this introduction, we provide a brief background on the two key points of the proposal to better facilitate review of the body and research accomplishments.

**Alternative Cleavage and Polyadenylation.** Recent large-scale Deep Sequencing analyses have found evidence for extensive global diversity of poly(A) site choice as an organism progresses through development or when quiescent T cells are activated using T cell antigen receptor (1-3). More importantly, these same changes in poly(A) site choice are observed as cells undergo transformation (4). The underlying trend in all of these studies is that an increased cell proliferation capacity (i.e. during tumorigenesis) is associated with a global shift from the dPAS to a pPAS, which predictably gives rise to more stable messages that are resistant to microRNA repression (Figure 1). This appears to be a general property of cancer cells as many different cancer types were tested for proximal PAS usage of a few select genes and many were found to have increased usage of these sites (4). *We hypothesized that genes vital to acquisition of androgen independence during prostate cancer tumor progression are regulated by APA. This was the goal of Aim 1.*



**Figure 1.** Shifting the polyA tail site to a proximal position (pPA) would result in a loss of microRNA inhibition of expression.

**H3K36 methylation.** Since the hypothesis of the histone code by Allis and colleagues, it has become clear that modifications of histones can have profound effects on gene expression. While there have been a myriad of studies on H3K36 in particular, the one most relevant to this research proposal was by Misteli and colleagues (5, 6). They observed that in prostate cancer cells that undergo epithelial to mesenchymal transition, H3K36 is trimethylated by Setd2. This trimethylation leads to recruitment of a reader protein that ultimately recruits a splicing regulator, PTB (Figure 2). There has been no link established, thus far, between histone methylation status and alternative cleavage and polyadenylation but it appears plausible. *We hypothesized that alterations in the activity of Setd2 may change H3K36 methylation status in nucleosomes near the 3' ends of genes regulating APA. This was the goal of Aim 2.*



**Figure 2.** Setd2 activity changes result in altered K36 trimethylation. Altered K36me3 results in differential recruitment of splicing factor PTB to cause different amounts of exon inclusion. (Figure from Wagner EJ & Carpenter, 2012)

## Body.

In this section, we describe both positive and negative results acquired during the one-year funding period. The two specific aims were conducted concurrently but, for the ease of reading and presentation, we present the results of Specific Aim 2 first.

**Specific Aim 2.** This goal of this aim was to determine the position of H3K36me at the 3' end of genes identified as undergoing APA and test the effect of Setd2 expression on APA.

Aim 2b. Modulate the expression level of Setd2 and determine the impact on APA using realtime PCR analysis.

To commence with this aim, we had to develop a real-time PCR assay on gene(s) undergoing APA that Setd2 may regulate. To this end, we devised an approach focusing on three genes well-known to undergo APA: cyclin D1, Timp2, and Dicer1. This approach would include designing amplicons specific to the mRNA using the distal polyA site choice as well as a common site to measure both distal and proximal. To do this, we tested a battery of PCR primers looking for pairs that would generate a single band on an agarose gel with a clean melting curve profile on a realtime PCR machine. As is seen in Figure 3, we successfully generated the amplicons for these three genes. Importantly, these amplicons measure the APA status of the endogenous genes so no reporter system was necessary.

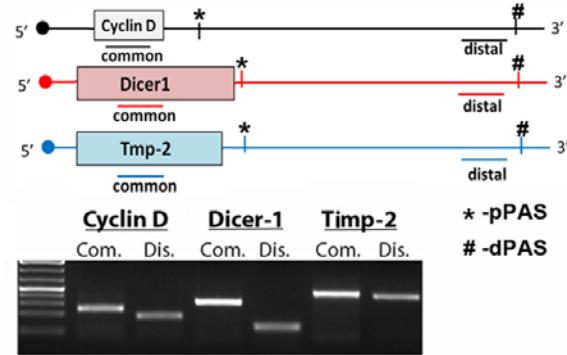


Figure 3. Amplicon design for three test genes known to have alternative cleavage and polyadenylation.

To downregulate the expression of Setd2, we designed and purchased siRNA specific to human Setd2. One of these siRNA was already validated to give >90% knockdown. Treatment of cells with these siRNA resulted in essentially no change in the level of APA usage for all three test genes (Figure 4, only cyclin D1 is shown). This is evidenced by the final column of the data demonstrating that the log difference in distal relative to common signal in both knockdowns is very similar to the control siRNA treated cells (C2).

	Cyclin D1					dCT (minus 7SK)				
	Common	Corr comn	Distal	7SK		Common	Distal	Diff	Log diff.	
C2	26.52	27.34	27.89	12.73	C2	14.61	15.16	0.55	1.464086	
C2	26.68	27.5	27.27	11.88	C2	15.62	15.39	-0.23	0.852635	
C2	26.69	27.51	27.63	11.84	C2	15.67	15.79	0.12	1.086735	
Set2 si1	26.87	27.69	26.77	12.84	Set2 si1	14.85	13.93	-0.92	0.528509	
Set2 si1	25.65	26.47	26.7	13.05	Set2 si1	13.42	13.65	0.23	1.172835	
Set2 si1	28.06	28.88	29.53	12.71	Set2 si1	16.17	16.82	0.65	1.569168	
Set2 si2	27.65	28.47	28.67	13.04	Set2 si2	15.43	15.63	0.2	1.148698	
Set2 si2	25.78	26.6	26.03	12.89	Set2 si2	13.71	13.14	-0.57	0.673617	
Set2 si2	26.86	27.68	27.31	12.03	Set2 si2	15.65	15.28	-0.37	0.773782	

Figure 4. Real-time PCR analysis of proximal versus distal choice of cyclin D1 mRNA. The values are Cts for each of the amplicons shown in figure 3 for cyclin D1. C2 represents the negative control siRNA

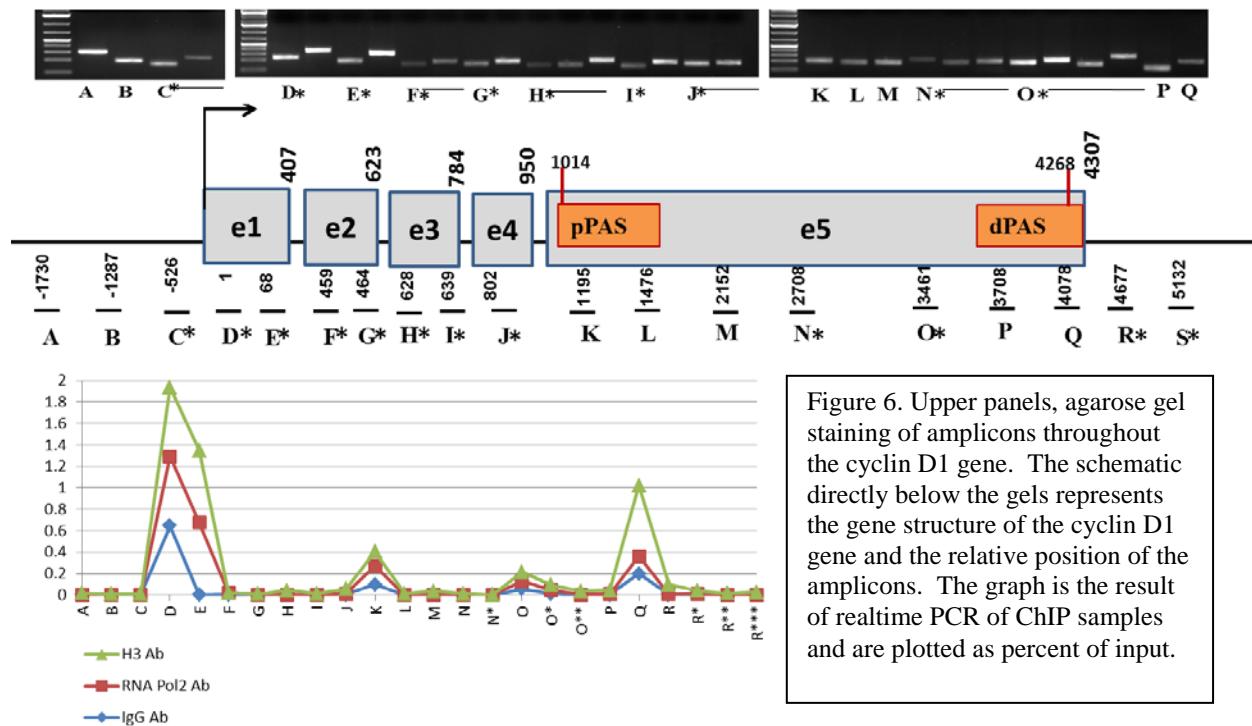
While we were confident of the technique and the ability of the siRNA to deplete Setd2, we sought to develop a positive control siRNA targeting a protein that modulated APA. Systematic testing of the 15 known polyA processing factors led us to focus our studies on CFIm25. Depletion of this protein from cells led to enhanced usage of the proximal polyA site, which is

what we anticipated, would happen after depletion of Setd2. What makes CFIm25 so interesting in prostate cancer is the observation that we made demonstrating that CFIm25 is downregulated in prostate cancer cell lines that are androgen-receptor negative (Figure 5). This result was followed up in Specific Aim 1 (see below).

Figure 5. Western Blot analysis from prostate cancer cell line lysates that are either androgen receptor positive or negative.

#### Aim 2a. Perform ChIP assays and real time PCR to map the position of H3K36me.

To conduct the ChIP assay, we designed a large set of amplicons throughout the cyclin D1 gene body, its promoter, and then downstream. The quality of the amplicons is of “ChIP-grade” as can be seen in Figure 6 (agarose gels). We performed ChIP assays using two control antibodies raised to RNA polymerase II (RNA Pol2) and histone H3. We observed significant histone H3 signal at the promoter and then again at the distal polyA site. The graph shown is representative of multiple experiments. We were not able to get an H3K36me3 antibody to work on this gene and trouble shooting was not continued given that the results from Specific Aim 2b did not suggest that Setd2 was functioning on this gene.



Specific Aim 1. The goal of this aim is to perform deep sequencing of mRNA isolated from different prostate cancer cell lines. Within this dataset we hope to identify novel instances of APA.

Aim 1a. Isolate total RNA, enrich for mRNA, perform 3' tagging on mRNA, and submit for deep sequencing.

This specific aim gave us the most trouble as, at the time of the grant submission, there was no developed protocol describing how to do 3' tagging to sequence polyA site tail junctions. Fortunately, a paper was published describing a new technology called “3PSeq” which was exactly the technique that we needed. We collaborated with a local Deep sequencing company, called LC Sciences, to develop 3PSeq. This technique is designed to partition down mRNA to just the few bases prior to the site of polyA tail addition as well as a few nucleotides into the polyA tail. As is seen in Figure 7, we were able to get sequences containing polyA tails at the expected positions; however, the level of sequence depth was not enough to make actual measurements.

sequence (+mRNA strand)	translated protein
TTTTAGTTGCTAAATAAAATGCCCTCTAAAACAAAAAA	STRAP
TTTTAGTTGCTAAATAAAATGCCCTCTAAAACAAAAAA	STRAP
CCTGCCCTCTCCCTGAATAAAAGAACAGCTTGACAGAAAA	RPL36
CCCTGTTCTGGACATTTCATATAAAATGGAATCACACAAAA	CBS
CCCTGTTCTGGACATTTCATATAAAATGGAATCACACAAAA	CBS
CTGTTCTGGACATTTCATATAAAATGGAATCACACAAAAAA	CBS
ATGGACCAGTCATAAAAGCCTCAGGCCCTCAAAAAAA	NDUFB7

Figure 7. Example sequencing reads from the 3PSeq experiment. The polyA sites are labeled in green and the polyA tail is in red.

In light of these results and the fact that we observed CFIm25 to be downregulated in AR-cell lines, we chose to conduct standard RNA-seq under conditions of CFIm25 RNAi versus control RNAi. The goal is then to use bioinformatics to filter only the 3' end reads to identify instances of APA that is regulated by CFIm25. Once these instances are identified, we would then use qRT-PCR to measure the APA status of these genes in AR+ versus AR- prostate cells. This turned out to be an excellent idea and constitutes one of the major findings during the funding period. We isolated RNA from control and CFIm25 knockdown cells and performed RNA-Seq. We achieved >115,000,000 reads for each sample, which is an outstanding dataset. Examples of the data are shown below where we determined the read density difference of cyclin D1, Dicer1, and Timp2 in control versus CFIm25 knockdown (Figure 8). In each case, it is evident that downregulation of CFIm25 leads to a reduction in read density in the 3' UTR proving that this factor regulates APA. Dr. Wei Li, of Baylor University, has designed software to analyze this data in a way that identifies all instances of 3'UTR shortening. We now have 2,181 candidate genes that are regulated by CFIm25. Gene ontology and pathway analysis uncovered the fact that the second most affected pathway of genes is androgen-signaling. We have requested a six month NCE in order to further analyze the list of genes undergoing APA after CFIm25 knockdown to test using qRT-PCR in cells that are androgen-dependent (AR+) versus androgen-independent (AR-).

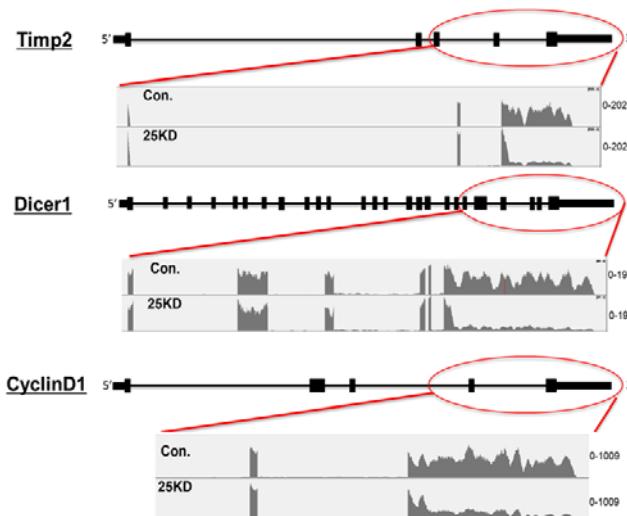


Figure 8. Read Densities generated from RNA-seq data from control siRNA treated cells versus CFIm25 knockdown cells.

Aim 1b. Analyze the dataset from deep sequencing and perform qRT-PCR validation of identified instances of APA.

-This will be performed during the NCE time period as we further analyze bioinformatically the data gathered from Aim 1a.

### **Key Research Accomplishments**

-Setd2 does not regulate genes known to undergo APA  
-CFIm25 is downregulated in androgen receptor negative PCa cells  
-CFIm25 regulates APA of genes involved in androgen signaling

### **Reportable outcomes:**

#### **Abstracts:**

1. Identification of factors that drive alternative cleavage and polyadenylation in cancer. Chioniso P, Masamha, Ann-Bin Shyu, **Wagner EJ**. Poster presented at the Eukaryotic mRNA Processing meeting (2011), CSHL, NY.

#### **Manuscripts:**

1. **\*Wagner EJ, \*Carpenter PB\***: Translating the Language of H3K36. *Nature Reviews in Molecular and Cellular Biology*. Jan 23;13(2):115-26, 2012 [PMID: 22266761] (\*co-corresponding authors)
2. Masamha CP, Zheng Y, Li W, Shyu AB, **Wagner EJ**. CFIm25 is a Global Repressor of Proximal PolyA Site Selection and Suppresses Growth (2012). Manuscript in preparation.

#### **Databases:**

Once we conclude our bioinformatics, we will generate a searchable database for CFIm25 targets that will be cited in the above manuscript and accessible through the Wagner laboratory webpage.

#### **Funding Applied for:**

NIH (1R01CA166274-01)-grant submitted in June 2011 and funded  
Title: Developing a HTS Assay for Inhibitors of Alternative Cleavage and Polyadenylation  
P.I.: Eric J. Wagner  
Project Period: 4/1/2012- 3/31/2014

## **Conclusions**

This proposal was to be carried out over the course of just one year and was meant to test a discrete hypothesis. While we were disappointed to determine that Setd2 was not regulating APA in a way that we could measure the identification of a factor that regulates APA that is altered in PCa cells is exciting. The androgen signaling pathway that is regulated by CFIIm25 is an important finding as very little is known about APA, thus this insight is novel. The questions that these findings have generated include: how is CFIIm25 downregulated in AR- cells and what impact does this have on the APA of PCa cells. These questions are new ones to the field and will act as a platform on which to base future, more substantial research proposals and programs.

## **References**

1. Sandberg R, Neilson JR, Sarma A, Sharp PA, & Burge CB (2008) Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. (Translated from eng) *Science* 320(5883):1643-1647 (in eng).
2. Mangone M, *et al.* (The landscape of *C. elegans* 3'UTRs. (Translated from eng) *Science* 329(5990):432-435 (in eng).
3. Jan CH, Friedman RC, Ruby JG, & Bartel DP (Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. (Translated from eng) *Nature* 469(7328):97-101 (in eng).
4. Mayr C & Bartel DP (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. (Translated from eng) *Cell* 138(4):673-684 (in eng).
5. Luco RF, *et al.* (2010) Regulation of alternative splicing by histone modifications. (Translated from eng) *Science* 327(5968):996-1000 (in eng).
6. Wagner EJ & Carpenter PB (2012) Understanding the language of Lys36 methylation at histone H3. (Translated from eng) *Nat Rev Mol Cell Biol* 13(2):115-126 (in eng).



# Understanding the language of Lys36 methylation at histone H3

Eric J. Wagner and Phillip B. Carpenter

**Abstract** | Histone side chains are post-translationally modified at multiple sites, including at Lys36 on histone H3 (H3K36). Several enzymes from yeast and humans, including the methyltransferases SET domain-containing 2 (Set2) and nuclear receptor SET domain-containing 1 (NSD1), respectively, alter the methylation status of H3K36, and significant progress has been made in understanding how they affect chromatin structure and function. Although H3K36 methylation is most commonly associated with the transcription of active euchromatin, it has also been implicated in diverse processes, including alternative splicing, dosage compensation and transcriptional repression, as well as DNA repair and recombination. Disrupted placement of methylated H3K36 within the chromatin landscape can lead to a range of human diseases, underscoring the importance of this modification.

#### Dosage compensation

The mechanism by which expression levels from sex chromosomes are adjusted. In mammalian systems, one copy of the X chromosome is silenced in the female. By contrast, in *Drosophila*, genes on the male X chromosome are expressed at twofold levels.

#### SET domain

(Suppressor of variegation 3-9, Enhancer of Zeste and Trithorax domain). A catalytic domain that uses S-adenosylmethionine to transfer methyl groups to substrates.

Department of Biochemistry and Molecular Biology, The University of Texas Medical School, Houston, Texas 77030, USA.

e-mails:

[Eric.J.Wagner@uth.tmc.edu](mailto:Eric.J.Wagner@uth.tmc.edu); [Phillip.B.Carpenter@uth.tmc.edu](mailto:Phillip.B.Carpenter@uth.tmc.edu)

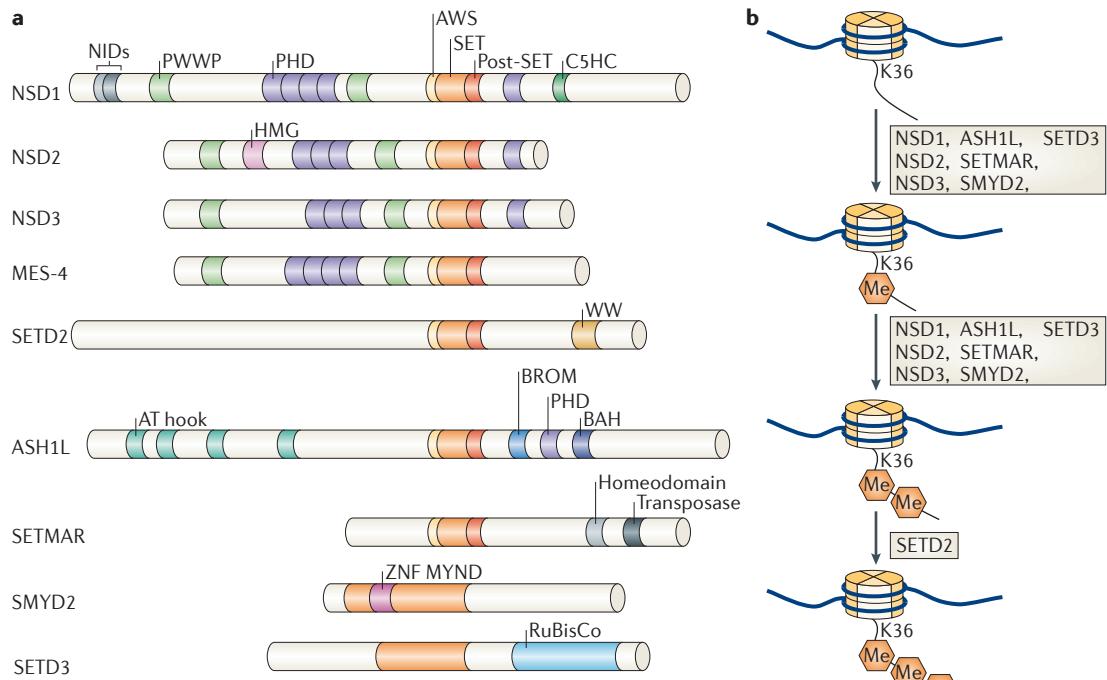
doi:10.1038/nrm3274

The packaging of DNA with basic histones and a vast array of additional factors mediates the formation of chromatin and thereby determines the outcome of virtually all of the DNA processes in eukaryotes. Acetylation and methylation of histones were originally identified in radiolabelling studies using cell extracts<sup>1,2</sup>. Today, enzymes catalysing numerous histone post-translational modifications or 'marks' have been identified; these include phosphorylation, various modes of methylation at Arg and Lys side chains, acetylation and ubiquitylation. These marks are thought to exist in dynamic combinations to generate a 'code' or 'language' that can enforce the regulatory features of chromatin during nearly all of the aspects of cellular metabolism: a given modification or permutation of modifications dictates a distinct biological output. Given the critical biological roles of chromatin and the numerous pathologies related to its misregulation, understanding the role of histone modifications is a paramount, yet complicated task. In this Review, we focus on our current understanding of a key modification, the methylation of Lys36 at histone H3 (H3K36). Widely described to be associated with active chromatin, H3K36 methylation has also been implicated in transcriptional repression, alternative splicing, dosage compensation, DNA replication and repair, DNA methylation and the transmission of the memory of gene expression from parents to offspring during development. We highlight the specificity, regulation and functional role of the enzymes that methylate H3K36, but we refer readers to informative reviews for a detailed discussion of the counteracting demethylases<sup>3–5</sup>.

#### Regulating the methylation of H3K36

Histone methyltransferase (HMTase) enzymes use S-adenosylmethionine to add methyl groups to specific histone Lys or Arg residues. To date, at least eight distinct mammalian enzymes have been described that methylate H3K36 *in vitro* and/or *in vivo* (FIG. 1). All of the H3K36-specific methyltransferases identified thus far have the catalytic SET domain in common, but they have varying preferences for Lys36 residues in different methylation states. In yeast, SET domain-containing 2 (Set2) performs all three methylation events at H3K36 (REF. 6), but in higher eukaryotes there is accumulating evidence that these events require a division of labour between the mono- and dimethylases and the SET2-type trimethylases (FIG. 1). Several of these enzymes also possess multiple chromatin-interacting domains, including those known to interact with methylated H3K36 itself (such as the PWWP domain) and with additional methylated histone residues (such as plant homeodomain fingers (PHD fingers))<sup>7–9</sup> (FIG. 1). Notably, SET2 proteins from multiple species have a carboxy-terminal domain that interacts with the large subunit of RNA polymerase II (RNAPII), which is known as RNAPII subunit B1 (RPB1)<sup>10</sup>.

**Determining HMTase specificity.** The most parsimonious scenario would predict that each H3K36 methyltransferase enzyme catalyses a transition between two distinct methylation states (for example, between unmethylated H3K36 and monomethylated H3K36 (H3K36me1)). However, this is not necessarily the case. There are many discrepancies as to the level of



**Figure 1 | Domain structures of enzymes that methylate H3K36.** **a** | A schematic of the enzymes that have been shown to promote the formation of methylated Lys36 on histone H3 (H3K36). With the exception of fly maternal-effect sterile 4 (MES-4), only human enzymes are shown. The SET domain is shown with its pre (AWS) and post domains; C5HC is a zinc-finger (ZNF) domain; WW domains are known to interact with Pro-rich peptides; the PWWP domain is known to interact with trimethylated H3K36; and the AT hook is a DNA-binding domain. All domain assignments were derived from Ensembl. **b** | Depiction of the transitions between the multiple H3K36 methylation states, highlighting the enzymes that have been shown to function in changing a given methylation state. ASH1L, ASH1-like; BAH, bromo-associated homology; BROM, bromodomain; HMG, high mobility group; MYND, myeloid, Nervy and DEAF-1 ZNF; NIDs, nuclear receptor interaction domains; NSD, nuclear receptor SET domain-containing; PHD, plant homeodomain; RuBisCo, ribulose-1,5-bisphosphate carboxylase oxygenase; SETD, SET domain-containing; SETMAR, SET domain and mariner transposase fusion gene-containing; SMYD2, SET and MYND domain-containing 2.

methylation (mono-, di- and tri-) imparted, as well as the residue (or residues) that is targeted, by the various H3K36-specific methyltransferases; these discrepancies may arise from multiple factors, including the nature of the substrate tested (for example, peptides and histones versus nucleosomes), the source of enzyme (for example, full length versus SET domain only) and the assay conditions themselves (for example, antibody specificity versus mass spectrometry). Physiologically relevant substrates, characterized using nucleosomes and/or loss-of-function experiments, have been determined for some H3K36 methyltransferases (TABLE 1); but for enzymes such as SETD3, the enzymatic activities reported so far are based on analyses of peptides and core histones<sup>11,12</sup>.

**PWPP domain**  
(Pro-Trp-Trp-Pro domain).  
A chromatin-interacting domain that has recently been shown to bind trimethylated Lys36 on histone H3.

**Plant homeodomain fingers**  
(PHD fingers). Motifs that bind zinc and have been shown to bind methylated residues.

consider this when interpreting phenotypes derived from loss-of-function experiments.

NSD1 has specific mono- and dimethylase activity for H3K36 (REFS 16–18), generating H3K36me1 and dimethylated H3K36 (H3K36me2), and there are also mixed reports as to whether it has specificity for H4K20 (REFS 14,19,20). Enzymatic assays using recombinant nucleosomes containing unmethylated H3K36 or a mimic of H3K36me1 (REF. 21) were shown to serve as specific substrates for NSD1 (REF. 16). However, when histone octamers were used as substrates, NSD1 was found to additionally methylate histone H4 as well as histones H2A and H2B, suggesting that nucleosomes contribute to H3K36 specificity. Structural data have also confirmed the specificity of NSD1 as a mono- and dimethylase that targets H3K36 (REF. 18). Intriguingly, this study showed that the substrate-binding channel of NSD1 is blocked by an autoinhibitory loop that resides between the SET and post-SET domains. As Lys36 resides near the core of the nucleosome, interactions between the enzyme and DNA may allosterically relieve inhibition by this loop, a result that is consistent with the preference of NSD enzymes for nucleosome substrates, as opposed to octamers. This further underscores the necessity for using physiologically relevant

Table 1 | Reported substrate specificities for enzymes that methylate H3K36

Enzyme*	Assay conditions	Histone substrates identified	Refs
NSD1 (KMT3B)	Nucleosomes	H3, H3K36, H3K36me1, H3K36me2, H4K20	14, 16–18
	Cell based and/or <i>in vivo</i>	H3K36me1, H3K36me2, H4K20me3	17–19, 78
NSD2 (WHSC1, MMSET)	Nucleosomes	H3, H3K36, H3K36me1, H3K36me2	16, 17, 19, 30
	Cell based and/or <i>in vivo</i>	H3K27me2 <sup>†</sup> , H3K36me2	17, 28, 36
NSD3 (WHSC1L1)	Nucleosomes	H3, H3K36	16
	Cell based and/or <i>in vivo</i>	H3K36me2	40
ASH1; ASH1L	Nucleosomes	H3, H3K36, H3K36me2, H3K36me3 <sup>§</sup>	41, 48, 49
	Cell based and/or <i>in vivo</i>	H3K4, H3K4me3	49
SMYD2	Core histones (octamers)	H3, H3K4, H3K36me2	42, 43
SETMAR (METNASE)	Core histones (octamers)	H3K4, H3K36me2	12
	Cell based and/or <i>in vivo</i>	H3K36me2	96
SETD2 (KMT3A, HYPB)	Nucleosomes	H3, H3K36me1, H3K36me2, H3K36me3	34, 64
	Cell based and/or <i>in vivo</i>	H3K36me3	25, 33, 34
SETD3	Core histones (octamers)	H3K4, H3K36	11

Table entries for the knockdown (cell based) or knockout (*in vivo*) analysis indicate a reported decrease in a given methylation state. ASH1, Absent, small and homeotic discs 1; ASH1L, ASH1-like; H, histone; NSD, nuclear receptor SET domain-containing; me1, monomethylated; me2, dimethylated; me3, trimethylated; SETD, SET domain-containing; SETMAR, SET domain and mariner transposase fusion gene-containing; SMYD2, SET and MYND domain-containing 2. \*Alternative names are provided in brackets. <sup>†</sup>RE-IIBP isoform of NSD2. <sup>§</sup>Gln2265Ala mutant of ASH1L.

substrates in order to properly characterize enzyme specificity towards H3K36.

NSD1 controls the levels of methylated H3K36 within and surrounding the body of the bone morphogenic protein 4 (*BMP4*) gene in human HCT116 colorectal cancer cells<sup>17</sup>. In the absence of NSD1, *BMP4* expression is significantly impaired and the levels of all three forms of methylated H3K36 are reduced within the body of the gene. These data could support a model in which NSD1 catalyses H3K36 trimethylation (H3K36me3), but a more likely scenario is that the loss of mono- and dimethylation at H3K36 also inhibits trimethylation by failing to provide another HMTase, SETD2 (also known as KMT3A and HYPB), with its substrate. This is an important distinction to make when considering which HMTases regulate the formation of a trimethylated residue.

The role of NSD1 as a mono- and dimethylase also seems to be conserved in other metazoans. Worms and flies each have one NSD1 orthologue called maternal-effect sterile 4 (MES-4)<sup>22,23</sup>. In *Caenorhabditis elegans*, MES-4 has been reported to be a dimethylase that is specific for H3K36 (REF. 22). Similarly to NSD1, *Drosophila melanogaster* MES-4 catalyses global mono- and dimethylation of H3K36 *in vivo*, but SET2, the fly orthologue of human SETD2, regulates all trimethylation at this site; again, MES-4 is probably required to provide the proper dimethylated H3K36 substrate for the SET2 enzyme.

#### Allosteric

Pertaining to the process by which a binding event at one site influences the activity of an enzyme or protein at a second, distant site. This may lead to activation or inhibition by cooperation between ligands, when a ligand bound at one site affects the affinity of another site for its ligand by inducing transitions between distinct conformational states.

In worms, mutation of the SETD2 orthologue *histone methyltransferase-like 1* (*met-1*) significantly reduced global H3K36me3 levels<sup>24</sup>. This supports the notion that the worm MES-4 and MET-1 proteins coordinate dimethylation and trimethylation, respectively<sup>24</sup>. These experiments are generally consistent with studies of mammalian NSD proteins and SETD2 (REFS 16, 25). By contrast, separate studies showed by immunostaining that worm embryos individually deficient in *mes-4* or *met-1* retain significant levels of H3K36me3, but embryos doubly deficient in *mes-4* and *met-1* are completely devoid of H3K36me3 (REFS 26, 27). This suggests that MES-4 collaborates with MET-1 to perform trimethylation. As each of these studies has used different antibodies, it will be important to validate these experiments using additional, antibody-independent techniques, such as mass spectrometry. Collectively, the combination of *in vitro* and *in vivo* data suggests that NSD1 and its orthologues probably catalyse the addition of either the monomethyl or dimethyl groups onto H3K36 and so indirectly regulate the levels of trimethylation by altering the availability of monomethyl and dimethyl substrates for the trimethylating enzymes.

**NSD2 as a mono- and dimethylase for H3K36.** Similarly to NSD1, NSD2, the product of Wolf–Hirschhorn syndrome candidate gene 1 (WHSC1; also known as MMSET), has been reported to target H3K36, as well as H4K20, H3K4 and H3K27 (REFS 16, 28–32). Here again, these discrepancies may lie, at least in part, with the nature of the substrates used in the assays. For example, NSD2 acts as a dimethylase towards H3K36 when presented with nucleosomes, but it preferentially dimethylates H4K44 when presented with octamers<sup>16</sup>. Interestingly, when short single-stranded and double-stranded DNA molecules that are notably smaller than the length that is necessary to generate a nucleosome are added to octamers, NSD2 preferentially dimethylates H3K36, a result that has been attributed to the ability of DNA to act as an allosteric effector of NSD2 (REF. 16). This is consistent with another study, in which mass spectrometry and immunoblotting against native and recombinant nucleosomes showed that NSD2, as well as its major splicing variant, RE-IIBP, acts as a mono- and dimethylase specific for H3K36 (REF. 30). By contrast, there have been reports that the NSD2 isoform RE-IIBP is an H3K27-specific HMTase<sup>28</sup> and that H3K36 trimethylation is lost in embryonic stem (ES) cells derived from NSD2-defective mice<sup>32</sup>. But the idea that NSD2 might act as a trimethylase would not be consistent with a previous study that identified SETD2 as the sole trimethylase in mammalian cells<sup>25</sup>. Furthermore, H3K36me3 levels are reduced when SETD2 is depleted, despite normal levels of H3K36me2 (REFS 33, 34). Regardless, these data are all consistent with a role for NSD2 in H3K36 methylation. This is also supported by the observation that, in ~15% of patients with multiple myeloma (a haematological malignancy that accounts for 1% of all cancers<sup>35</sup>), NSD2 is overexpressed as a result of a translocation with the immunoglobulin locus ( $t(4;14)^+$ ) and the global levels of H3K36me2 are increased<sup>36</sup>.

**Possible specificity of NSD2 for H4K20.** Despite strong evidence linking NSD2 to H3K36me, it may also act as an H4K20-specific methyltransferase during the cellular response to DNA double-strand breaks (DSBs)<sup>37,38</sup>. NSD2 was isolated in an RNA interference screen for genes involved in resistance to hydroxyurea, a DNA-replication inhibitor<sup>38</sup>. Consistently, NSD2-defective cells are sensitive to DNA damage and NSD2 localizes to sites of DNA damage. Moreover, NSD2 is phosphorylated at Ser102 by the kinase ataxia telangiectasia mutated (ATM) upon damage<sup>37</sup>, and this mediates the recruitment of NSD1 to sites of DNA damage, where it has been suggested to dimethylate H4K20, which is an important event in recruiting the DNA damage response regulator p53-binding protein 1 (53BP1)<sup>39</sup>. Although one could envision that NSD2 phosphorylation and damaged DNA may both alter the catalytic specificity of NSD2 through allosterism, it should be noted that recombinant NSD2 methylates H4 in *in vitro* assays with H4 peptides<sup>37</sup>. Thus, it is possible that NSD2 does have some specificity for H4 in addition to H3 even in the absence of DNA damage.

**The Set2 trimethylase.** In yeast, Set2 is non-essential, performs all three methylation reactions at H3K36, and couples H3K36me3 with transcriptional elongation through an interaction with Rpb1, the large subunit of RNAPII<sup>10</sup>. The human orthologue of Set2, SETD2, also interacts with RNAPII during elongation, but it is an essential protein and has been identified as a huntingtin-interacting protein. The role of SETD2 in Huntington's disease is obscure. SETD2 is a trimethylase<sup>25,33,34</sup>, and some reports have indicated that, *in vitro*, recombinant human SETD2 can add all three methyl groups to H3K36 (REFS 32,34). By contrast, *in vivo* knockdown experiments targeting SETD2 revealed reduced levels of H3K36me3 only<sup>34</sup>. This underscores the importance of using multiple approaches to determine enzyme specificity and also raises an important point about enzyme processivity. Does a trimethylase obligatorily require a dimethyl substrate or can it processively place all three groups on a Lys substrate? SETD2 associates with heterogeneous nuclear ribonucleoprotein L (hnRNPL), and hnRNPL knockdown analyses show decreased levels of H3K36me3 but not H3K36me1 or H3K36me2 (REF. 34). Thus, the ability to act as a processive enzyme could depend on the presence and regulation of distinct cofactors, as well as on the presence of a previously methylated substrate.

**Heterogeneous nuclear ribonucleoprotein L (hnRNPL).** A member of a highly abundant family of RNA-binding proteins known to associate with newly synthesized pre-mRNA.

**Pro isomerization**  
Peptide bonds to Pro residues can exist in either a *cis* or a *trans* state, and the transitions can be catalysed by prolyl isomerase enzymes.

**Anti-silencing function 1 (Asf1).** A histone chaperone protein that associates with newly synthesized histones and is involved in chromatin synthesis.

to dimethylation, but other substrate specificities have also been described for these enzymes<sup>12,41–43</sup>. However, the assays for SETD3 specificity were performed with peptides<sup>11</sup>, and it will be important to confirm this enzyme specificity using nucleosome-based assays<sup>16,44</sup>.

In addition to targeting H3K36, SMYD2 methylates H3K4 and non-histone substrates, such as p53 and retinoblastoma (RB)<sup>45,46</sup>. The *in vitro* methylase activity of SMYD2 towards H3K4, but not towards H3K36, depends on its ability to bind heat shock protein 90 $\alpha$  (HSP90 $\alpha$ )<sup>42</sup>. Although the expression of SMYD2 appears to be largely confined to the brain and heart, its conditional deletion from mouse cardiomyocytes showed that it is dispensable for heart development and, surprisingly, its deletion had no effect on the global levels of methylated H3K36 (REF. 47). This may suggest that the physiological target for SMYD2 is a non-histone substrate, providing an additional example of how the interpretation of knockout phenotypes for HMTases must be carefully considered.

Rigorous experiments using recombinant nucleosomes have demonstrated that ASH1L is a dimethylase that is specific for H3K36 (REFS 41,48). However, ASH1L has also been reported to target H3K4 and to localize within the transcribed regions of active target genes<sup>49</sup>; and *in vivo* levels of H3K4me3 were reduced in the absence of ASH1L. Additional studies have shown that ASH1L also targets H3K36 and H4K20 (REFS 44,46). Thus, it remains to be seen what the full substrate repertoire of ASH1L is. Notably, similarly to NSD1, ASH1L has an autoinhibitory loop that resides between the SET and post-SET domains, which is a feature that has been proposed as a hallmark feature of H3K36-specific enzymes<sup>41,48</sup>.

**Additional factors that influence H3K36 methylation.** Several mechanisms have been described for how H3K36 methylation levels are regulated at a given locus (REFS 50–52). In addition to cofactors such as HSP90 $\alpha$  and hnRNPL, Pro isomerization of histone H3 and additional histone modifications on both H3 and H4 have been shown to regulate the total levels of H3K36 methylation by Set2, although it is not clear how this occurs. Additional studies in yeast have shown that cyclins (encoded by bypass UAS requirement 1 (*BUR1*) and *BUR2*) are required for H3K36me3 (REF. 53). Moreover, the yeast histone chaperone anti-silencing function 1 (Asf1), the RNAPII kinase C-terminal domain (CTD) kinase 1 (Ctk1) and the elongation factor Spt6 (a subunit of the FACT (facilitates chromatin transcription) complex) all regulate the levels of H3K36 trimethylation but not dimethylation<sup>54,55</sup>. Other elongation factors, such as polymerase-associated factor 1 (Paf1), have also been shown to alter the levels of H3K36 methylation<sup>56</sup>. Large cells 1 (Lge1), a factor associated with the ubiquitin pathway, was also identified in a genome-wide screen for candidate genes that specifically regulate H3K36 methylation<sup>57</sup>. This study also confirmed the role of Paf1 in regulating H3K36me3: mutations in Paf1 disrupted the levels of H3K36me3, but not those of H3K36me2 (REF. 56). Collectively, these studies suggest

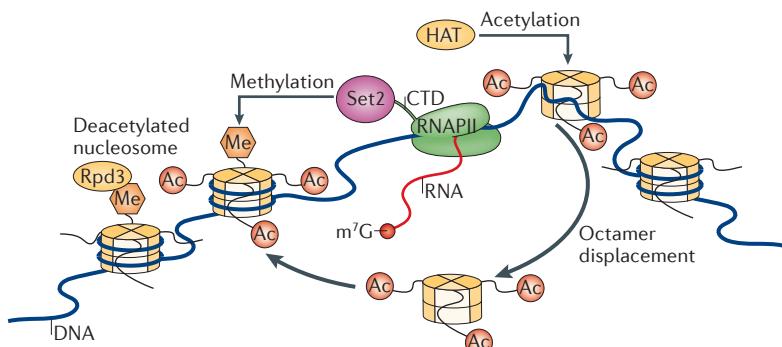


Figure 2 | H3K36me3-dependent prevention of aberrant transcription in yeast.

Actively transcribing RNA polymerase II (RNAPII) displaces acetylated nucleosomes. These evicted histones are reincorporated into nucleosomes and chromatin behind the polymerase. As SET domain-containing 2 (Set2) binds RNAPII, this promotes the trimethylation of Lys36 on histone H3 (H3K36me3) in the newly incorporated nucleosomes. H3K36me3 serves as a 'mark' that the reduced potassium dependency 3 (Rpd3) deacetylase complex binds; this complex facilitates local nucleosome deacetylation, preventing aberrant, spurious transcription in the wake of RNAPII progression through a region. CTD, carboxy-terminal domain; HAT, histone acetylase; m<sup>7</sup>G, 7-methylguanosine.

an emerging theme: that HMTase cofactors may alter substrate specificity on a gene-to-gene basis, providing an additional layer of regulation for their activity at distinct targets.

### Roles of H3K36 methylation in gene expression

Methylation of H3K36 has been observed at multiple stages of RNA biosynthesis and is not always associated with gene activation. Rather, its ultimate output and where it fits into gene expression pathways are functions of multiple variables: where in the gene body the methylated H3K36 mark is placed, when H3K36 is methylated and what reader protein binds to this modification. Further complicating these variables is the fact that the degree of methylation can result in different biological outcomes.

**H3K36 methylation and transcriptional activation.** Numerous studies in multiple systems support a role for H3K36 methylation in transcriptional activation. It has been observed that, in general, there is a progressive shift from monomethylation to trimethylation of H3K36 between the promoters and the 3' ends of genes<sup>58</sup>. However, in zebrafish, somatic cells also show a bias for H3K36me3 in the 3' end of actively transcribed genes, but this mark is curiously present in the 5' promoter regions of quiescent genes that are developmentally regulated during spermatogenesis<sup>59</sup>. The functional significance of this is unknown, and it will be important to determine whether promoter-proximal H3K36me3-containing nucleosomes can recruit repressive histone deacetylases to promote gene silencing.

Multiple lines of evidence in budding yeast have showed that Set2, a non-essential enzyme that is responsible for all three forms of methylated H3K36, is coupled to transcriptional elongation. Deletions in yeast Set2 cause sensitivity to the elongation inhibitor 6-azauracil

and phenocopy mutations in other known elongation factors<sup>10</sup>. However, it remains possible that this phenotype may arise from indirect effects on nucleotide metabolism. Nevertheless, Set2 associates with the hyperphosphorylated form of RNAPII and deposits the trimethyl group onto H3K36 during elongation<sup>10,60–62</sup> in various systems, including yeast and humans. This interaction is regulated by phosphorylated residues in the CTD of Rpb1, the large subunit of RNAPII: the Ser2-phosphorylated form of the CTD is indicative of elongation and the Ser5-phosphorylated form is characteristic of a paused polymerase at a promoter<sup>63</sup>. Indeed, human SET2 proteins also appear to bind RNAPII and to target H3K36 (REFS 34,64).

One well-established function of Set2 in yeast is the prevention of aberrant transcriptional initiation within coding sequences (FIG. 2). Set2 catalyses H3K36 methylation co-transcriptionally and recruits the reduced potassium dependency 3 small (Rpd3S) deacetylase complex here<sup>65–67</sup> through association of the chromodomain-containing Rpd3S subunit ESA1-associated factor 3 (Eaf3) with H3K36me3; this enforces a deacetylated chromatin state in the wake of transcribing RNAPII<sup>65,68</sup>. Rpd3S preferentially associates with histones containing H3K36me2 and H3K36me3 but not H3K36me1 (REF. 57). The PHD finger of the Rpd3S subunit Rco1 can also cooperate with the Eaf3 chromodomain to promote these interactions<sup>69</sup>. This indicates that H3K36me2 and H3K36me3 might act redundantly in the Set2–Rpd3S pathway.

The role of histone Lys methylation in the maintenance of a repressive chromatin environment during transcriptional elongation is also used in humans but may be independent of acetylation<sup>70</sup>. In this case, Lys demethylase 2 (LSD2; also known as KDM1B) demethylates methylated H3K4 in the intragenic regions of active target genes. LSD2 resides in complexes with NSD3, which acts as a mono- and dimethylase that is specific for H3K36 (REF. 16). Additionally, LSD2 binds the H3K9-specific HMTase G9A (also known as EHMT2). Consistent with this, the LSD2 complex has robust HMTase activity towards H3K9 and, to a lesser extent, H3K36 (REF. 70). Moreover, in chromatin immunoprecipitation followed by microarray (ChIP-chip) experiments, depletion of G9A alters both H3K9 methylation levels and transcriptional regulation at LSD2 targets. In light of these observations and the fact that LSD2 resides in complexes with elongation factors such as cyclin T1 (CCNT1; also known as PTEFb) and the Ser2-phosphorylated form of RNAPII, it appears that the methylation-dependent enforcement of a repressive chromatin state through H3K9 and possibly H3K36 methylation at actively transcribed genes is evolutionarily conserved. Additionally, it seems that NSD3 can participate in transcriptional elongation<sup>70</sup>. Interestingly, the LSD2 complex contains multiple proteins with PWWP domains, including NSD3. Originally identified in NSD2, PWWP domains also bind H3K36me3 to act as reader proteins<sup>7,8,71</sup>. In particular, the PWWP domain of DNA methyltransferase 3A (DNMT3A) binds H3K36me3 and methylates nearby DNA,

#### Reader protein

A protein that recognizes and binds post-translational modifications on histones.

#### Chromodomain

A domain of ~50 residues that has been shown to bind methylated residues.

demonstrating a link between H3K36me3 and DNA methylation that could represent a new mechanism for H3K36 methylation-mediated repression of gene expression.

**H3K36 methylation in dosage compensation.** Coordination between methylation and acetylation also occurs during dosage compensation in *D. melanogaster*<sup>72</sup>. To compensate for having only one X chromosome, male flies use the Male-specific lethal (MSL) complex to upregulate the expression of X-linked genes by a factor of two in order to achieve the same level of transcription as females, which have two X chromosomes. This may occur through a two-step model requiring H3K36me3 and increased elongation<sup>73–75</sup>. The MSL complex is initially recruited to X-linked genes through a GA-rich recognition site and subsequently facilitates transcriptional elongation through an enhancement of RNAPII activity. Next, the MSL3 subunit (an orthologue of yeast Eaf3) is thought to promote spreading of the upregulated state through interactions that involve the chromodomain of MSL3 and H3K36me3 (REF. 73). Additional active marks, such as H4K16 acetylation, also affect X-linked gene expression, requiring the Males absent on the first (MOF) subunit of MSL. Thus, a collection of activating marks together achieves chromosome-wide gene expression<sup>74</sup>. Although interactions between chromodomain proteins and H3K36me3 marks regulate acetylation in both yeast and *D. melanogaster*, they have opposite effects: transcription is repressed in yeast but upregulated in flies. This clearly illustrates that the function of a given histone modification, such as H3K36me3, is context-dependent and is much more complicated than a simple code.

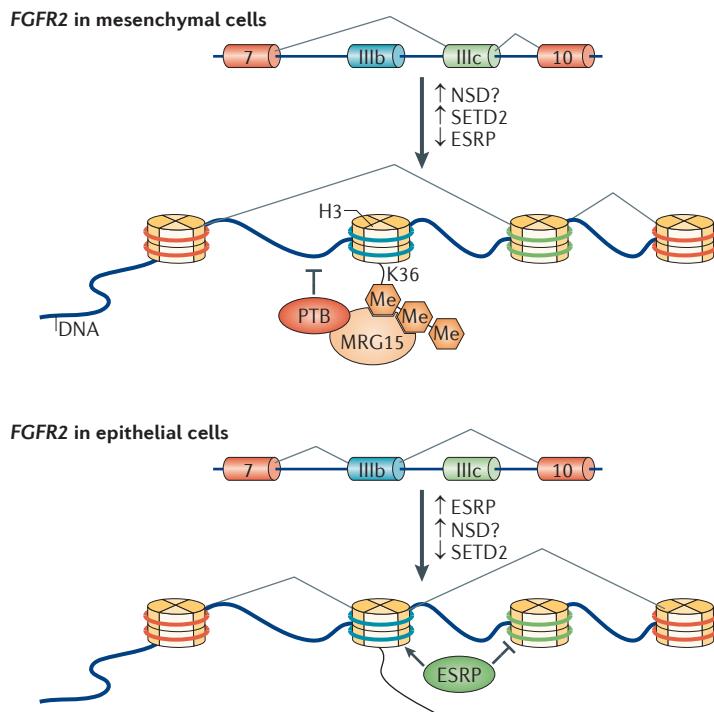
**NSD enzymes and transcriptional initiation.** NSD1 regulation of H3K36 methylation appears to affect transcriptional initiation<sup>17</sup>. For example, NSD1 binds upstream of the promoter of *BMP4* and regulates the levels of H3K36me1, H3K36me2 and H3K36me3 within the body of the gene. This seems to be required for the recruitment of RNAPII to the *BMP4* promoter, an observation that links NSD1 to initiation events through RNAPII. NSD3 can also bind promoters and then influence the levels of H3K36 methylation within the body of a gene<sup>40</sup>. For example, the extraterminal domains of bromodomain-containing 4 (BRD4) specifically recruit NSD3 to the promoter region of the gene encoding *CCND1* and the decapping enzyme DCPS, and loss of either BRD4 or NSD3 reduces the levels of H3K36me3, particularly within the body of the gene. This is consistent with the NSD1-dependent regulation of *BMP4*, during which promoter-proximal bound NSD1 regulates H3K36me3 levels within the body of the gene<sup>17</sup>. Although NSD3 localization is biased towards the promoter of *CCND1*, it also can be found, albeit at reduced levels, at the 3' end of the gene<sup>40</sup>, suggesting that NSD3 participates in initiation and elongation events. Moreover, NSD3 levels were found to be higher in the coding regions than in the promoter of the proto-oncogene Ser/Thr kinase *PIM2* (REF. 40).

Thus, NSD3 appears to reside in multiple complexes (including LSD2 and BRD4 complexes) and can localize to either promoter or internal regions to promote H3K36 methylation and thereby influence initiation and elongation processes. This is in apparent contrast to NSD1, which has been reported to localize primarily near the 5' ends of its targets<sup>17</sup>; it will be important to confirm this using high-resolution ChIP followed by sequencing (ChIP-seq) analysis. However, previous ChIP-seq data show that NSD2 is enriched near transcription start sites<sup>30</sup>.

**H3K36 methylation in transcriptional repression.** Fascinating clues have emerged regarding the relationship between methylation at H3K36, transcriptional repression and worm development<sup>26,27</sup>. In worms, maternally provided MES-4 is essential for germ cell viability and is involved in X-chromosome silencing in the germline<sup>26,27</sup>. Early primordial germ cells (PGCs) do not engage in transcription as they lack the active hyperphosphorylated CTD form of RNAPII. But, in the absence of *mes-4*, active RNAPII persists, indicating that MES-4 is important for the establishment and/or maintenance of transcriptional repression in late stage PGCs<sup>26,27</sup>. Loss of MES-4 also reduces the levels of H3K36me3, independently of RNAPII function. Thus, MES-4 maintains the H3K36me3 state in germline precursors independently of transcription — an observation that has not been made in other organisms so far.

Beyond H3K36 methylation-mediated regulation of spurious initiation, H3K36 methylation-mediated repression has also been reported. Yeast Set2 represses transcription from a *lacZ* reporter and, in Set2-deletion mutants, the basal levels of *GAL4* transcription are increased<sup>76,77</sup>; however, in these studies, Set2 is artificially recruited to the promoter, so it will be important to confirm whether this reflects the true nature of Set2 activity. Initial studies with tethered reporters in mammalian cells have implicated NSD1 in both transcriptional activation and repression<sup>13</sup>, although its targets have not yet been determined. In addition, an RNA interference study has shown that NSD1 represses the expression of the homeobox regulator *MEIS1* in neuronal cell lines<sup>19</sup>. This repressive event is likely to be direct, as NSD1 localizes to the promoter of *MEIS1* (REF. 78). Furthermore, NSD2 collaborates with the cardiac-specific factor NKX2-5 to repress targets such as platelet-derived growth factor- $\alpha$  (PDGFR $\alpha$ ), probably through modulating H3K36 methylation levels<sup>32</sup>. Thus, H3K36 methylation appears to act as both an activating and inhibitory signal, and the overall biological readout might depend on the context of additional surrounding marks and their corresponding reader proteins.

**H3K36 methylation and exon definition.** There has been increasing support for the idea that biased nucleosome positioning favours exonic over intronic sequences<sup>79–81</sup>. Such a bias seems to be accompanied by distinct histone modifications, including H3K36me3. This should not be surprising, as the 147-base-pair length of a DNA fragment associated with a histone



**Figure 3 | H3K36me3 influences alternative splicing in a cell-type specific manner.** The fibroblast growth factor receptor 2 (FGFR2) locus that undergoes alternative splicing consists of two mutually exclusive exons, IIIb and IIIc, which are located between the constitutive exons 7 and 10. Mesenchymal stem cells favour the inclusion of exon IIIc and achieve this by repressing splicing of exon IIIb. Nucleosomes present near exon IIIb contain the SET domain-containing 2 (SETD2)-dependent trimethylated Lys36 on histone H3 (H3K36me3) 'mark' and its reader protein MORF-related gene 15 (MRG15). MRG15 also interacts with polypyrimidine tract-binding protein (PTB), a known repressor of exon inclusion, and this may be the mechanism by which the methylated H3K36 mark can influence splicing at this locus. In epithelial cells, FGFR2 expresses exon IIIb but excludes exon IIIc. Epithelial splicing regulatory protein (ESRP) is expressed and stimulates the inclusion of exon IIIb; reduced levels of H3K36me3 present at this exon, possibly as a result of lower SETD2 levels, allow its derepression. The role of dimethylases, such as the proteins of the nuclear receptor SET domain-containing (NSD) family, in this process has yet to be determined but these enzymes could also influence H3K36 methylation here.

**Polypyrimidine tract-binding protein (PTB).** A protein that has been implicated as an antagonist of exon definition, the action of which results in the repression of exon inclusion.

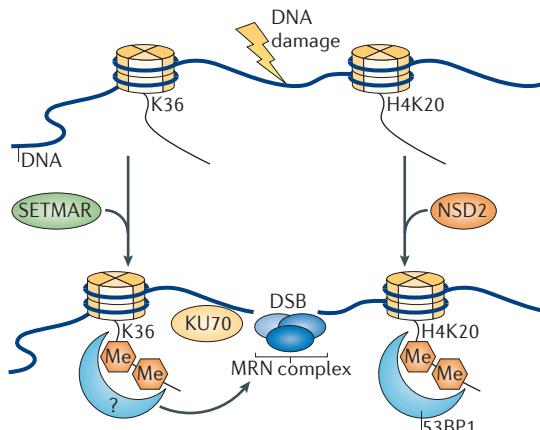
**Fibroblast growth factor receptor 2 (FGFR2).** A membrane-bound receptor that undergoes alternative splicing and is subject to regulation by methylation of Lys36 on histone H3.

**Epithelial splicing regulatory protein (ESRP).** An alternative splicing factor that is enriched in epithelial tissues and is responsible for enforcing specific exon inclusion.

are displaced further downstream. The known association between Set2 and the RNAPII CTD may also further explain the particular increase in H3K36me3 signatures seen at downstream exons<sup>61,82</sup>. The implication of nucleosome enrichment and the increase in H3K36me3 modifications is twofold. First, nucleosomes probably act as intrinsic pause sites for elongating RNAPII which could alter splice site recognition and, hence, change exon inclusion. Others have found that the introduction of pause sites within minigenes can increase the inclusion of alternatively spliced exons<sup>83,84</sup>. Furthermore, expression of a 'slow' mutant RNAPII in *D. melanogaster* results in different inclusion patterns of the exons within the *Ultrabithorax* mRNA<sup>85</sup>. A second possibility, which is not mutually exclusive with effects on RNAPII pausing, is that the H3K36me3 modification relays a specific signal to the splicing machinery to alter how it defines exons, leading to the specific inclusion or exclusion of particular exons.

Although the global analyses of H3K36me3 positioning in various genomes provide compelling evidence that this modification affects splicing, functional evidence beyond this has been lacking. However, an interesting connection has been made between SETD2, the reader protein MORF-related gene 15 (MRG15; which contains a chromodomain) and polypyrimidine tract-binding protein (PTB), the last of which is a known antagonist of exon definition that affects splicing of fibroblast growth factor receptor 2 (FGFR2) pre-mRNA<sup>86,87</sup>. FGFR2 contains two mutually exclusive exons (IIIb and IIIc) that encode a region within the extracellular immunoglobulin-like domain and are each responsible for receptor binding to a unique range of FGFs<sup>88</sup>. Exon IIIb is included in epithelial cells through the action of epithelial splicing regulatory protein (ESRP), whereas exon IIIc is included in cells of mesenchymal origin<sup>89</sup> (FIG. 3). Furthermore, the splicing pattern switches from exon IIIb to exon IIIc inclusion as prostate epithelial cells become androgen-independent, an important factor in metastasis. Analysis of nucleosome modifications throughout FGFR2 show that H3K36me3 is specifically enriched within exon IIIb and is restricted to mesenchymal cells, which exclude this exon. This H3K36me3 modification is recognized by MRG15, which also interacts with PTB. Thus, by recruiting PTB to its target exon, these interactions position PTB to bind to its intronic splicing silencer sites, which flank the repressed exon as they emerge from the transcribing RNAPII complex (FIG. 3). PTB repression of this exon can be alleviated by downregulating either MRG15 or SETD2. This regulatory module also exists at other alternatively spliced exons, with a bias towards exons that contain weaker PTB-binding sites<sup>86</sup>. What remains to be seen is how two distinct cell types achieve this differential methylation of H3K36 within nucleosomes at alternatively spliced exons in order to regulate splicing.

This example of FGFR2 control exemplifies how the methylation status of H3K36 can affect splicing. However, this crosstalk is bidirectional: mutations in splice sites that abrogate intron removal of  $\beta$ -globin reporter genes cause a shift in H3K36me3 signatures



**Figure 4 | Model for H3K36 methylation at sites of DNA damage.** Upon the generation of DNA damage, SET domain and mariner transposase fusion gene-containing (SETMAR) dimethylates Lys36 on histone H3 (H3K36) near sites of DNA double-strand breaks (DSBs), possibly by recruiting currently undefined reader proteins that facilitate the binding of KU70 and the MRE11-RAD50-NBS1 (MRN) complex to facilitate DNA repair. How methylation at H3K36 is coordinated with other chromatin modifications that are known to participate in DNA repair (such as, p53-binding protein 1 (53BP1) binding to dimethylated H4K20 (H4K20me2)) is unknown. NSD2, nuclear receptor SET domain-containing 2.

towards the 3' region of genes<sup>90</sup>. Moreover, others have observed that H3K36me3 signatures are markedly lower in genes without introns<sup>91</sup>. In both cases, inhibition of splicing with pharmacological agents causes a rapid and global redistribution of H3K36me3 throughout the genome and reduces SETD2 recruitment to these genes<sup>90,91</sup>. Although it is unclear how the spliceosome regulates H3K36 methylation, these studies suggest that it is a general phenomenon.

**H3K36 methylation in DNA replication, recombination and repair.** As replication origins initiate at distinct times in S phase, the decision to fire at any particular origin depends on a number of factors, including the transcriptional status of nearby genes. Origins of replication are bound by the origin recognition complex (ORC), which recruits various factors, including CDC6, minichromosome maintenance (MCM) proteins and CDC45, before loading of the DNA polymerase. Studies in budding yeast have revealed a role for H3K36 methylation in regulating replication origin firing, as deletions in *SET2* cause a delay in Cdc45 loading at origins<sup>92</sup>. H3K36 methylation has also been linked to DNA replication checkpoint control in fission yeast<sup>93</sup>. Furthermore, yeast mutants in the FACT elongation complex are sensitive to the replication inhibitor hydroxyurea, and this sensitivity can be suppressed by mutations in *SET2* (REF. 94). How H3K36 methylation influences these S phase events is unknown, but it could keep chromatin in an active state, perhaps through recruiting specific reader proteins that can influence DNA replication control.

#### Non-homologous end joining

(NHEJ). The main pathway that is used primarily in the G1 phase of the cell cycle to repair chromosomal DNA double-strand breaks in somatic cells. In contrast to homologous-recombination repair, NHEJ is error-prone because it leads to the joining of heterologous ends.

Histone methylation is also critical for the maintenance of genomic stability, including the biological response to DSBs<sup>39,95</sup>. For example, H3K36me2 has been implicated in the repair response to DSBs through the non-homologous end-joining (NHEJ) pathway<sup>96</sup>. Here, the methylase SETMAR, a SET domain-containing enzyme previously reported to affect NHEJ as well as DNA replication, was shown to directly mediate H3K36me2 formation near sites of DSBs<sup>96</sup> (FIG. 4). SETMAR-dependent formation of H3K36me2 was shown to enhance the rate of association of the DNA repair factors KU70 and Nijmegen breakage syndrome 1 (NBS1; also known as nibrin) near DSBs<sup>96</sup>. However, it is not yet clear how methylated H3K36 acts at the sites of DSBs, particularly with respect to which proteins may respond to the presence of this mark, or what relationship this mark might have with other known histone modifications that are important for DSB repair. Most notably, this may include crosstalk with dimethylation of H4K20, an event that recruits the DNA damage response regulator 53BP1 to sites of damage<sup>99</sup>. It is also possible that methylated H3K36 in this context mediates its effects by antagonizing transcriptional elongation, as RNAPII is displaced in response to DNA damage<sup>97</sup>.

#### H3K36 methylation: links with disease

In higher eukaryotes, defects in the genes that maintain the levels of H3K36 methylation cause developmental defects and disease<sup>32,33,98</sup> (TABLE 2). Defects in *SETD2* are causal for sporadic clear renal cell carcinoma, a disease that is marked by the loss of the short arm of chromosome 3 (REF. 99). *SETD2* has also been hypothesized to be a tumour suppressor in breast cancer<sup>100</sup>. Defects in each member of the NSD family have been implicated in multiple diseases and cancer types<sup>31,78,101–106</sup>. For example, haploinsufficiency in *NSD1* is causal for Sotos overgrowth syndrome<sup>98</sup>. *NSD1* has also been implicated in breast, lung and prostate cancer, acute myeloid leukaemia (AML) and refractory anaemia<sup>19,31,78,101–103,107–112</sup>. *NSD1*-defective mice display an embryonic-lethal phenotype: the embryos initiate mesoderm formation but fail to complete gastrulation, apparently as a result of apoptosis<sup>14</sup>. *NSD2*-defective mice die shortly after birth and display features that are consistent with Wolf–Hirschhorn syndrome<sup>32</sup>, including facial abnormalities and cardiac defects. Thus, *NSD1*-knockout mice and *NSD2*-knockout mice have different phenotypes, despite the fact that both proteins catalyse the formation of H3K36me2. No knockout-mouse phenotype has yet been reported for *NSD3*.

Each NSD family member behaves as an oncogene in multiple cancers<sup>35,78,105,112</sup>. Translocations in *NSD1* or *NSD3* lead to the development of AML<sup>105,112</sup>. Moreover, *NSD1*, an androgen receptor co-regulator<sup>113</sup>, has been identified as a candidate gene capable of discriminating between malignant and non-malignant prostate tissue<sup>101</sup>. In addition, silencing of *NSD1* has also been shown to cause sensitivity to the oestrogen receptor antagonist tamoxifen<sup>114</sup>, but how these two observations are linked to the ability of *NSD1* to bind and regulate either the androgen or oestrogen receptors is unknown. Overexpression of *NSD2* has also been implicated in multiple myeloma and additional cancers, such as neuroblastoma<sup>102,115</sup>, and

Table 2 | H3K36 methyltransferases and their roles in diseases

Gene	Diseases	Molecular defects	Phenotypes	Refs
NSD1	Sotos syndrome	Haploinsufficiency, point mutations, deletions, translocations	Macrocephaly, hypertolerism, cognitive and/or motor skill deficiencies	98
NSD1	Myelodysplastic syndrome	Translocations	Anaemia, cytopenia	111
NSD1	Cancers: AML, prostate, neuroblastoma, breast	Overexpression, gene silencing, translocations	Numerous tumour types	19,78,101, 110,112
NSD2	Wolf–Hirschhorn syndrome	Deletions	Learning difficulties, microcephaly, heart defects	32
NSD2	Multiple myeloma	t(4;14) <sup>+</sup> translocation	Renal failure, anaemia, bone lesions	30,115
NSD3	Breast cancer	Gene amplification at 8p11	Solid tumours	107
NSD3	AML	Translocations	Leukaemic cells in bone marrow	105
NSD3	Myelodysplastic syndrome	Translocations	Anaemia, cytopenia	106
SETD2	Renal cell carcinoma	Deletions, missense mutations	Haematuria, flank pain	99

The table lists the known diseases resulting from alterations in the genes encoding enzymes that are specific for Lys36 on histone H3 (H3K36). AML, acute myeloid leukaemia; NSD, nuclear receptor SET domain-containing; SETD2, SET domain-containing 2.

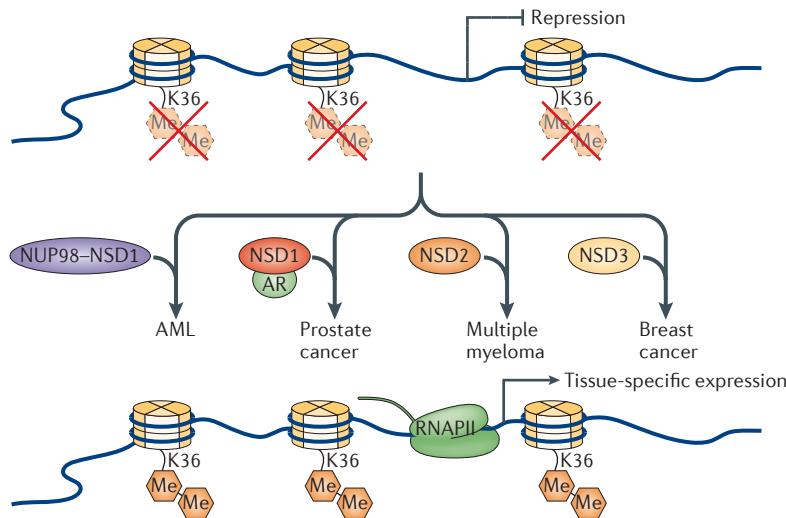
overexpression of NSD3 via gene amplification occurs in about 15% of breast cancers<sup>107</sup> and correlates with poor prognosis. Therefore, NSD proteins are instrumental in the development of cancer, but their mechanisms of action in this context are only beginning to be elucidated. One possible mechanism underlying NSD-mediated disease is that these proteins enforce H3K36 methylation patterns to change gene expression. This is discussed below in the context of well-known NSD-dependent cancers.

Through overexpression or their incorporation into fusion proteins via translocations, NSD proteins act as potent oncogenes<sup>30,78</sup>. For example, fusion proteins between NSD1 and the nucleoporin NUP98 are well established as being causal for a subset of AML<sup>112</sup>, and they regulate homeobox (*Hox*) gene expression in a mouse model of AML<sup>78</sup>. In the case of AML, NSD1–NUP98 fusions enforce an H3K36 methylation signal that contributes to the inappropriate activation of *Hox* genes during development. These increased H3K36 methylation levels seem to be antagonistic to the repressive methylated H3K27 marks that normally silence *Hox* gene expression, leading to reduced H3K27 methylation levels and gene activation. As a consequence, unscheduled cell proliferation and the generation of AML are observed in the mouse model<sup>78</sup>.

Overexpression of NSD2 via a translocation between chromosome 4 and chromosome 14 alters the global profile of H3K36 methylation in KMS11 multiple myeloma cells<sup>30,36</sup>. Moreover, the catalytic activity of NSD2 is required for tumorigenesis in a mouse xenograft model<sup>30</sup>. And loss of NSD2 function (through either short hairpin RNA-mediated depletion or homologous recombination) globally decreases H3K36me2 and H3K36me3 levels but has no effect on H4K20 or H3K4 methylation<sup>36</sup>. It is, however, accompanied by a concomitant decrease in the global levels of H3K27me2 and H3K27me3, and this downregulation results in the expression of target genes that would normally be quiescent. This suggests

that the overexpression of NSD2 alters transcriptional programming by promoting the formation of H3K36me2 and thereby changing the chromatin structure. Indeed, ChIP-seq analysis has precisely mapped the distribution of H3K36me2 in KMS11 multiple myeloma cells overexpressing NSD2 (REF. 30). In cells that have one normal allele of NSD2, the H3K36me2 signal revealed by ChIP-seq was preferentially enriched in intragenic regions, which is consistent with previous ChIP-chip results in flies<sup>23</sup>. However, overexpression of NSD2 disperses an increased H3K36me2 signal throughout the genome, including intergenic regions. Cells expressing one normal allele of NSD2 have a modest H3K36me2 signal in promoter regions, followed by a peak at transcriptional start sites, and then the signal decays downstream to the 3' end<sup>30</sup>, and a high signal corresponds with higher transcriptional activation. By contrast, cells overexpressing NSD2 have little variance in H3K36me2 signal intensity across the average gene and the correlation between signal levels and transcription is lost. Instead, as some genes are much more sensitive to H3K36me2 levels than others, this aberrant enrichment of the H3K36me2 signal appears to trigger the expression of quiescent oncogenes, including transforming growth factor alpha (TGFA), MET, p21 activated kinase 1 (PAK1) and *RRAS2* (related RAS viral (*r-ras*) oncogene homologue)<sup>30</sup>.

H3K36 and H3K27 methylation may be mutually exclusive modifications<sup>44</sup>. Nucleosomes pre-methylated at H3K27 are largely refractory to enzymatic catalysis at H3K36 and vice versa. As overexpression of NSD3 is causal for the generation of breast cancer through 8p11–12 amplifications<sup>107,109</sup>, one exciting possibility is that global gene expression profiles are drastically altered through a switch from H3K27 to H3K36 methylation. When considering this and the fact that antagonism between H3K36 and H3K27 methylation has also been observed in worms<sup>22</sup>, one might anticipate that this could be a general mechanism for the oncogenic behaviour observed



**Figure 5 | NSD proteins can act as oncogene proteins.** Through overexpression and/or translocation events that result in the fusion of nuclear receptor SET domain-containing (NSD) proteins with other proteins, such as the nucleoporin NUP98, NSD proteins can be aberrantly recruited to target loci in various tissues. As a consequence, global levels of dimethylated Lys36 on histone H3 (H3K36me2) increase and are sufficient to activate inappropriate transcription, which contributes to cancer development. In some cases, the increased levels of H3K36me2 are expected to inversely correlate with trimethylated H3K27 (H3K27me3; not shown), altering the balance between competitive activating and repressive 'marks'. As a consequence, multiple gene sets that are sensitive to the levels of H3K36me2 are turned on, a causal event in oncogenesis. AML, acute myeloid leukaemia; AR, androgen receptor; RNApII, RNA polymerase II.

for enzymes that promote H3K36 or H3K27 methylation (FIG. 5). But NSD proteins are not solely restricted to acting as oncogene proteins; they can also act as tumour suppressors, as a lack of NSD1 expression is observed in neuroblastomas<sup>19</sup>.

## Conclusions

Although widely perceived as an activating modification, H3K36 methylation also functions in transcriptional repression as well as in processes such as DNA repair. Thus, H3K36 methylation function extends beyond transcription and is likely to be relevant in a wide range of DNA-based processes. So far, much effort has gone into understanding the nature of the enzymes and their substrate specificities, but little is known about how the enzymes that mediate H3K36 methylation are localized and regulated beyond the well-characterized model of yeast Set2 and its relationship with RNApII. Given that NSD1 binds various nuclear receptors, it is clear that this H3K36-specific methyltransferase, and probably others, operates in distinct tissues and perhaps in separate complexes. Similarly, NSD2 interacts with different transcription factors in mouse ES cells versus heart cells from embryonic day 12.5 (REF 32). Thus, elucidating the distinct complexes that these enzymes form in different cell types and at different stages of development will be important. As mammalian cells possess multiple, non-redundant enzymes that methylate H3K36 to varying degrees, the roles of this modification are expected to be more complex than in yeast. Despite this, it is apparent that a role for methylated H3K36 in coordinating crosstalk between acetylation and methylation is conserved between flies and higher eukaryotes<sup>116</sup>. This crosstalk results in a given H3K36 modification that is interpreted by the reader protein in the context of neighbouring histone modifications, potentially changing its meaning. Undoubtedly, this 'combinatorial histone language' will be the subject of future studies of methylated H3K36 that will expand its scope of importance to the chromatin landscape.

1. Allfrey, V. G., Faulkner, R. & Mirsky, A. E. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci. USA* **51**, 786–794 (1964).
2. Allfrey, V. G. & Mirsky, A. E. Structural modifications of histones and their possible role in the regulation of RNA Synthesis. *Science* **144**, 559 (1964).
3. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nature Rev. Genet.* **7**, 715–727 (2006).
4. Varier, R. A. & Timmers, H. T. Histone lysine methylation and demethylation pathways in cancer. *Biochim. Biophys. Acta* **1815**, 75–89 (2011).
5. Rotili, D. & Mai, A. Targeting histone demethylases: a new avenue for the fight against cancer. *Genes Cancer* **2**, 663–679 (2011).
6. Lee, J. S. & Shilatifard, A. A site to remember: H3K36 methylation a mark for histone deacetylation. *Mutat. Res.* **618**, 130–134 (2007).
7. Dhayalan, A. *et al.* The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* **285**, 26114–26120 (2010).
8. Vezzoli, A. *et al.* Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nature Struct. Mol. Biol.* **17**, 617–619 (2010).
9. Baker, L. A., Allis, C. D. & Wang, G. G. PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. *Mutat. Res.* **647**, 3–12 (2008).
10. Kizer, K. O. *et al.* A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol. Cell. Biol.* **25**, 3305–3316 (2005).
11. Eom, G. H. *et al.* Histone methyltransferase SETD3 regulates muscle differentiation. *J. Biol. Chem.* **286**, 34733–34742 (2011).
12. Lee, S. H. *et al.* The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair. *Proc. Natl. Acad. Sci. USA* **102**, 18075–18080 (2005).
13. Huang, N. *et al.* Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. *EMBO J.* **17**, 3398–3412 (1998).
14. Rayasam, G. V. *et al.* NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J.* **22**, 3153–3163 (2003).
15. Lu, T. *et al.* Regulation of NF- $\kappa$ B by NSD1/FBXL11-dependent reversible lysine methylation of p65. *Proc. Natl. Acad. Sci. USA* **107**, 46–51 (2010).
16. Li, Y. *et al.* The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J. Biol. Chem.* **284**, 34283–34295 (2009). Shows, using methyl-Lys analogues in nucleosomal substrates, that proteins of the NSD family are highly specific for the dimethylation of nucleosomal H3K36.
17. Lucio-Eterovic, A. K. *et al.* Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proc. Natl. Acad. Sci. USA* **107**, 16952–16957 (2010).
18. Qiao, Q. *et al.* The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *J. Biol. Chem.* **286**, 8361–8368 (2011). Reveals that NSD1 activation and catalysis occur through an interaction with the nucleosomes, which causes a conformational change to relieve inhibition by the autoinhibitory loop.
19. Berdasco, M. *et al.* Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. *Proc. Natl. Acad. Sci. USA* **106**, 21830–21835 (2009).
20. Yang, H. *et al.* Preferential dimethylation of histone H4 lysine 20 by Suv4-20. *J. Biol. Chem.* **283**, 12085–12092 (2008).
21. Simon, M. D. *et al.* The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128**, 1003–1012 (2007). Introduces the technique of generating methyl-Lys analogues as a tool to understand histone methylation and enzyme specificity.
22. Bender, L. B. *et al.* MES-4: an autosome-associated histone methyltransferase that participates in silencing the X chromosomes in the *C. elegans* germ line. *Development* **133**, 3907–3917 (2006).
23. Bell, O. *et al.* Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in *Drosophila*. *EMBO J.* **26**, 4974–4984 (2007).
24. Andersen, E. C. & Horvitz, H. R. Two *C. elegans* histone methyltransferases repress *lin-3* EGF transcription to inhibit vulval development. *Development* **134**, 2991–2999 (2007).
25. Edmunds, J. W., Mahadevan, L. C. & Clayton, A. L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* **27**, 406–420 (2008).

26. Furuhashi, H. *et al.* Trans-generational epigenetic regulation of *C. elegans* primordial germ cells. *Epigenetics Chromatin* **3**, 15 (2010).

27. Rechtsteiner, A. *et al.* The histone H3K36 methyltransferase MES-4 acts epigenetically to transmit the memory of germline gene expression to progeny. *PLoS Genet.* **6** pii: e1001091 (2010).

28. Kim, J. Y. *et al.* Multiple-myeloma-related WHSC1/MMSET isoform RE-IIBP is a histone methyltransferase with transcriptional repression activity. *Mol. Cell. Biol.* **28**, 2023–2034 (2008).

29. Kim, S. M. *et al.* Characterization of a novel WHSC1-associated SET domain protein with H3K4 and H3K27 methyltransferase activity. *Biochem. Biophys. Res. Commun.* **345**, 318–323 (2006).

30. Kuo, A. J. *et al.* NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol. Cell* **44**, 609–620 (2011). **ChIP-seq data with NSD2 in multiple myeloma cells reveal that H3K36me2 generates a chromatin landscape that is permissive for tumorigenesis by activating oncogenes that are sensitive to H3K36me2 levels.**

31. Marango, J. *et al.* The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor. *Blood* **111**, 3145–3154 (2008).

32. Nimura, K. *et al.* A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf–Hirschhorn syndrome. *Nature* **460**, 287–291 (2009).

33. Hu, M. *et al.* Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. *Proc. Natl. Acad. Sci. USA* **107**, 2956–2961 (2010).

34. Yuan, W. *et al.* Heterogeneous nuclear ribonucleoprotein L is a subunit of human KMT3a/ Set2 complex required for H3 Lys-36 trimethylation activity *in vivo*. *J. Biol. Chem.* **284**, 15701–15707 (2009).

35. Chesi, M. *et al.* The t(4;14) translocation in myeloma dysregulates both *FGFR3* and a novel gene, *MMSET*, resulting in IgH/MMSET hybrid transcripts. *Blood* **92**, 3025–3034 (1998).

36. Martinez-Garcia, E. *et al.* The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells. *Blood* **117**, 211–220 (2011).

37. Pei, H. *et al.* MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. *Nature* **470**, 124–128 (2011).

38. Hajdu, I., Ciccia, A., Lewis, S. M. & Elledge, S. J. Wolf–Hirschhorn syndrome candidate 1 is involved in the cellular response to DNA damage. *Proc. Natl. Acad. Sci. USA* **108**, 13130–13134 (2011).

39. Botuyan, M. V. *et al.* Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361–1373 (2006).

40. Rahman, S. *et al.* The Brd4 extraterminal domain confers transcriptional activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol. Cell. Biol.* **31**, 2641–2652 (2011).

41. Tanaka, Y., Kataegiri, Z., Kawahashi, K., Kioussis, D. & Kitajima, S. Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene* **397**, 161–168 (2007).

42. Abu-Farha, M. *et al.* The tale of two domains: proteomics and genomics analysis of SMYD2, a new histone methyltransferase. *Mol. Cell Proteomics* **7**, 560–572 (2008).

43. Brown, M. A., Sims, R. J. 3rd, Gottlieb, P. D. & Tucker, P. W. Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. *Mol. Cancer* **5**, 26 (2006).

44. Yuan, W. *et al.* H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* **286**, 7983–7989 (2011).

45. Huang, J. *et al.* Repression of p53 activity by Smyd2-mediated methylation. *Nature* **444**, 629–632 (2006).

46. Saddic, L. A. *et al.* Methylation of the retinoblastoma tumor suppressor by SMYD2. *J. Biol. Chem.* **285**, 37733–37740 (2010).

47. Diehl, F. *et al.* Cardiac deletion of Smyd2 is dispensable for mouse heart development. *PLoS ONE* **5**, e9748 (2010).

48. An, S., Yeo, K. J., Jeon, Y. H. & Song, J. J. Crystal structure of the human histone methyltransferase ASH1L catalytic domain and its implications for the regulatory mechanism. *J. Biol. Chem.* **286**, 8369–8374 (2011). **Shows that ASH1L is a dimethylase that is highly specific for nucleosomal H3K36 and that histone tails pre-methylated at H3K36 are refractory to methylation at H3K27 and vice versa.**

49. Gregory, G. D. *et al.* Mammalian ASH1L is a histone methyltransferase that occupies the transcribed region of active genes. *Mol. Cell. Biol.* **27**, 8466–8479 (2007).

50. Du, H. N., Fingerman, I. M. & Briggs, S. D. Histone H3 K36 methylation is mediated by a trans-histone methylation pathway involving an interaction between Set2 and histone H4. *Genes Dev.* **22**, 2786–2798 (2008).

51. Nelson, C. J., Santos-Rosa, H. & Kouzarides, T. Proline isomerization of histone H3 regulates lysine methylation and gene expression. *Cell* **126**, 905–916 (2006).

52. Psathas, J. N., Zheng, S., Tan, S. & Reese, J. C. Set2-dependent K36 methylation is regulated by novel intratal interactions within H3. *Mol. Cell. Biol.* **29**, 6413–6426 (2009).

53. Chu, Y., Sutton, A., Sternglanz, R. & Prelich, G. The BUR1 cyclin-dependent protein kinase is required for the normal pattern of histone methylation by SET2. *Mol. Cell. Biol.* **26**, 3029–3038 (2006).

54. Lin, L. J., Minard, L. V., Johnston, G. C., Singer, R. A. & Schultz, M. C. Asf1 can promote trimethylation of H3 K36 by Set2. *Mol. Cell. Biol.* **30**, 1116–1129 (2010).

55. Youdell, M. L. *et al.* Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. *Mol. Cell. Biol.* **28**, 4915–4926 (2008).

56. Chu, Y., Simic, R., Warner, M. H., Arndt, K. M. & Prelich, G. Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. *EMBO J.* **26**, 4646–4656 (2007).

57. Li, B. *et al.* Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3S histone deacetylase complex and to repress spurious transcription. *J. Biol. Chem.* **284**, 7970–7976 (2009).

58. Bannister, A. J. *et al.* Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J. Biol. Chem.* **280**, 17732–17736 (2005).

59. Wu, S. F., Zhang, H. & Cairns, B. R. Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. *Genome Res.* **21**, 578–589 (2011).

60. Li, B., Howe, L., Anderson, S., Yates, J. R. 3rd & Workman, J. L. The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* **278**, 8897–8903 (2003).

61. Li, J., Moazed, D. & Gygi, S. P. Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. *J. Biol. Chem.* **277**, 49383–49388 (2002).

62. Xiao, T. *et al.* Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev.* **17**, 654–663 (2003).

63. Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20**, 2922–2936 (2006).

64. Sun, X. J. *et al.* Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *J. Biol. Chem.* **280**, 35261–35271 (2005).

65. Carrozza, M. J. *et al.* Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581–592 (2005).

66. Joshi, A. A. & Struhl, K. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol. Cell* **20**, 971–978 (2005).

67. Keogh, M. C. *et al.* Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**, 593–605 (2005).

68. Kaplan, C. D., Laprade, L. & Winston, F. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* **301**, 1096–1099 (2003).

69. Li, B. *et al.* Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science* **316**, 1050–1054 (2007).

70. Fang, R. *et al.* Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation. *Mol. Cell* **39**, 222–233 (2010).

71. Bartke, T. *et al.* Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470–484 (2010).

72. Bell, O. *et al.* Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in *Drosophila melanogaster*. *Mol. Cell. Biol.* **28**, 3401–3409 (2008).

73. Alekseyenko, A. A. *et al.* A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. *Cell* **134**, 599–609 (2008).

74. Gelbart, M. E., Larschan, E., Peng, S., Park, P. J. & Kuroda, M. I. *Drosophila* MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nature Struct. Mol. Biol.* **16**, 825–832 (2009).

75. Larschan, E. *et al.* X chromosome dosage compensation via enhanced transcriptional elongation in *Drosophila*. *Nature* **471**, 115–118 (2011).

76. Strahl, B. D. *et al.* Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol. Cell. Biol.* **22**, 1298–1306 (2002).

77. Landry, J. *et al.* Set2-catalyzed methylation of histone H3 represses basal expression of GAL4 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **23**, 5972–5978 (2003).

78. Wang, G. G., Cai, L., Pasillas, M. P. & Kamps, M. P. NUP98–NSD1 links H3K36 methylation to *Hox*-A gene activation and leukaemogenesis. *Nature Cell Biol.* **9**, 804–812 (2007). **Shows, using a mouse model system, that a NUP98–NSD1 translocation enforces an activating H3K36 methylation profile at a subset of *Hoxa* genes, causing uncontrolled cell proliferation and tumorigenesis.**

79. Kolasinska-Zwierz, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genet.* **41**, 376–381 (2009).

80. Schwartz, S., Meshorer, E. & Ast, G. Chromatin organization marks exon–intron structure. *Nature Struct. Mol. Biol.* **16**, 990–995 (2009).

81. Spies, N., Nielsen, C. B., Padgett, R. A. & Burge, C. B. Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* **36**, 245–254 (2009).

82. Krogan, N. J. *et al.* Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.* **23**, 4207–4218 (2003).

83. Roberts, G. C., Gooding, C. & Smith, C. W. Smooth muscle alternative splicing induced in fibroblasts by heterologous expression of a regulatory gene. *EMBO J.* **15**, 6301–6310 (1996).

84. Robson-Dixon, N. D. & Garcia-Blanco, M. A. MAZ elements alter transcription elongation and silencing of the fibroblast growth factor receptor 2 exon IIIb. *J. Biol. Chem.* **279**, 29075–29084 (2004).

85. de la Mata, M. *et al.* A slow RNA polymerase II affects alternative splicing *in vivo*. *Mol. Cell* **12**, 525–532 (2003).

86. Luco, R. F. *et al.* Regulation of alternative splicing by histone modifications. *Science* **327**, 996–1000 (2010). **Demonstrates that H3K36me3 modification is capable of regulating alternative splicing events, notably those regulated by PTB, through the direct recruitment of splicing regulators to nucleosomes.**

87. Wagner, E. J. & Garcia-Blanco, M. A. Polyprymidine tract binding protein antagonizes exon definition. *Mol. Cell. Biol.* **21**, 3281–3288 (2001).

88. Carstens, R. P., McKeehan, W. L. & Garcia-Blanco, M. A. An intronic sequence element mediates both activation and repression of rat fibroblast growth factor receptor 2 pre-mRNA splicing. *Mol. Cell. Biol.* **18**, 2205–2217 (1998).

89. Warzecha, C. C., Sato, T. K., Nabet, B., Hogenesch, J. B. & Carstens, R. P. ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol. Cell* **33**, 591–601 (2009).

90. Kim, S., Kim, H., Fong, N., Erickson, B. & Bentley, D. L. Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc. Natl. Acad. Sci. USA* **108**, 13564–13569 (2011).

91. de Almeida, S. F. *et al.* Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nature Struct. Mol. Biol.* **18**, 977–983 (2011).

92. Pryde, F. *et al.* H3 k36 methylation helps determine the timing of Cdc45 association with replication origins. *PLoS ONE* **4**, e5882 (2009).

93. Kim, H. S., Rhee, D. K. & Jang, Y. K. Methylation of histone H3 lysine 9 and lysine 36 are functionally linked to DNA replication checkpoint control in fission yeast. *Biochem. Biophys. Res. Commun.* **368**, 419–425 (2008).

94. Biswas, D. *et al.* A role for Chd1 and Set2 in negatively regulating DNA replication in *Saccharomyces cerevisiae*. *Genetics* **178**, 649–659 (2008).

95. Sanders, S. L. *et al.* Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell* **119**, 603–614 (2004).

96. Fnu, S. *et al.* Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. *Proc. Natl Acad. Sci. USA* **108**, 540–545 (2011).

97. Chou, D. M. *et al.* A chromatin localization screen reveals poly (ADP-ribose)-regulated recruitment of the repressive Polycomb and NuRD complexes to sites of DNA damage. *Proc. Natl Acad. Sci. USA* **107**, 18475–18480 (2010).

98. Kurotaki, N. *et al.* Haploinsufficiency of *NSD1* causes Sotos syndrome. *Nature Genet.* **30**, 365–366 (2002).

99. Duns, G. *et al.* Histone methyltransferase gene *SETD2* is a novel tumor suppressor gene in clear cell renal cell carcinoma. *Cancer Res.* **70**, 4287–4291 (2010).

100. Newbold, R. F. & Mokbel, K. Evidence for a tumour suppressor function of *SETD2* in human breast cancer: a new hypothesis. *Anticancer Res.* **30**, 3309–3311 (2010).

101. Bianco-Miotto, T. *et al.* Global levels of specific histone modifications and an epigenetic gene signature predict prostate cancer progression and development. *Cancer Epidemiol. Biomarkers Prev.* **19**, 2611–2622 (2010).

102. Hudlebusch, H. R. *et al.* The histone methyltransferase and putative oncoprotein mmset is overexpressed in a large variety of human tumors. *Clin. Cancer Res.* **17**, 2919–2933 (2011).

103. Hudlebusch, H. R. *et al.* MMSET is highly expressed and associated with aggressiveness in neuroblastoma. *Cancer Res.* **71**, 1226–4235 (2011).

104. Kassambara, A., Klein, B. & Moreaux, J. MMSET is overexpressed in cancers: link with tumor aggressiveness. *Biochem. Biophys. Res. Commun.* **379**, 840–845 (2009).

105. Rosati, R. *et al.* *NUP98* is fused to the *NSD3* gene in acute myeloid leukemia associated with t(8;11) (p11.2;p15). *Blood* **99**, 3857–3860 (2002).

106. Taketani, T. *et al.* *NUP98-NSD3* fusion gene in radiation-associated myelodysplastic syndrome with t(8;11)(p11;p15) and expression pattern of NSD family genes. *Cancer Genet. Cytogenet.* **190**, 108–112 (2009).

107. Angrand, P. O. *et al.* *NSD3*, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines. *Genomics* **74**, 79–88 (2001).

108. Job, B. *et al.* Genomic aberrations in lung adenocarcinoma in never smokers. *PLoS ONE* **5**, e15145 (2010).

109. Zhang, H. *et al.* Forkhead transcription factor foxq1 promotes epithelial–mesenchymal transition and breast cancer metastasis. *Cancer Res.* **71**, 1292–1301 (2011).

110. Zhao, Q. *et al.* Transcriptome-guided characterization of genomic rearrangements in a breast cancer cell line. *Proc. Natl Acad. Sci. USA* **106**, 1886–1891 (2009).

111. La Starza, R. *et al.* Cryptic insertion producing two *NUP98/NSD1* chimeric transcripts in adult refractory anemia with an excess of blasts. *Genes Chromosomes Cancer* **41**, 395–399 (2004).

112. Jaju, R. J. *et al.* A novel gene, *NSD1*, is fused to *NUP98* in the t(5;11)(q35;p15.5) in *de novo* childhood acute myeloid leukemia. *Blood* **98**, 1264–1267 (2001).

113. Wang, X. *et al.* Identification and characterization of a novel androgen receptor coregulator ARA267- $\alpha$  in prostate cancer cells. *J. Biol. Chem.* **276**, 40417–40423 (2001).

114. Mendes-Pereira, A. M. *et al.* Genome-wide functional screen identifies a compendium of genes affecting sensitivity to tamoxifen. *Proc. Natl Acad. Sci. USA* 11 Apr 2011 (doi:10.1073/pnas.1018872108).

115. Keats, J. J. *et al.* Overexpression of transcripts originating from the *MMSET* locus characterizes all t(4;14)(p16;q32)-positive multiple myeloma patients. *Blood* **105**, 4060–4069 (2005).

116. Morris, S. A. *et al.* Identification of histone H3 lysine 36 acetylation as a highly conserved histone modification. *J. Biol. Chem.* **282**, 7632–7640 (2007).

#### Acknowledgements

We apologize to those researchers whose work could not be cited owing to space constraints. We thank B. Martinez and K. Eterovic for assistance with the figures, as well as S. Ercan, P. Masamha and A. Sataluri for critical comments regarding the manuscript. Work in the laboratory of P.B.C. has been supported in part by the Robert A. Welch Foundation (AU-1569). Work in the laboratory of E.J.W. is supported by the US National Institutes of Health (5R00GM080447).

#### Competing interests statement

The authors declare no competing financial interests.

#### FURTHER INFORMATION

Ensembl: <http://www.ensembl.org>

#### FURTHER INFORMATION

Eric J. Wagner's homepage: <http://www.uth.tmc.edu/bmb/faculty/eric-wagner.html>

Phillip B. Carpenter's homepage: <http://www.uth.tmc.edu/bmb/faculty/Carpenter/Carpenter.html>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF